DOCUMENT RESUME

ED 243 585 PS 014 278

AUTHOR Ellett, Chad D.
TITLE Issues Related to the Evaluation of Program Implementation in Follow Through.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 12 Mar 81
NOTE 34p.; Paper presented at the National Institute of Education's Follow Through Planning Conference (Philadelphia, PA, February 10-11, 1981).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Data Analysis; Early Childhood Education; Models; *Program Descriptions; *Program Evaluation; *Program Implementation; *Research Methodology
IDENTIFIERS *Causal Models; Planned Variation; *Project Follow Through

ABSTRACT

This paper provides a rationale for studies of educational program implementation and offers a causal model for program evaluation emphasizing pupil-related variables that intervene between program implementation and outcome variables. The rationale is followed by a discussion of implementation as an evaluation construct. Process, independent variable, and dependent variable perspectives are considered. Current models for measuring program implementation are described. While none is comprehensive, each model can be adapted to selected aspects of Follow Through. The question of where implementation fits on "the measurement continuum" is explored, and the view is expressed that implementation as a program evaluation construct allows for "weaker" and more global measurement indices. In contrast, an educational "treatment" construct requires greater measurement precision and more detailed specification of program variables. Both approaches are thought to have analogues relevant to designing a comprehensive evaluation/experimentation plan for future studies in Follow Through. The succeeding discussion emphasizes the need for a systematic and comprehensive description of key program components as a precondition to developing evaluation instruments. Examples of "critical dimensions" of two Follow Through program components, associated performance indicators, and scaled descriptors, capable of being scored, are provided. In conclusion, deficiencies of the comparative, longitudinal approach to evaluating Follow Through are pointed out and the alternate, causal model is described. (RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

PS

ED243585

ISSUES RELATED TO THE EVALUATION OF PROGRAM IMPLEMENTATION IN FOLLOW THROUGH

Chad D. Ellett
College of Education
University of Georgia

Paper prepared for a conference sponsored by the National Institute of Education entitled Documentation of School Improvement Efforts: Some Technical Issues and Future Research Agenda

Learning Research and Development Center
University of Pittsburgh

March 12, 1981

PS 014278

This paper was written under contract with the National Institute of Education. The opinions expressed are those of the author and do not necessarily represent the positions or policies of NIE.

ISSUES RELATED TO THE EVALUATION OF PROGRAM IMPLEMENTATION IN FOLLOW THROUGH

In a 1978 paper, Herb Walberg reiterated that two perennial questions in education are: "What are the ends of education?" and "Do the educational means, that is, the manipulations of the environment, justify the ends?" These two questions seem to provide a succinct conception of the focus of this conference on school improvement efforts. From a philosophical perspective, these generic questions and their many answers are inextricably bound to matters of human values, ethics, and morality. The disciplines of educational and psychological measurement and educational evaluation raise additional key points: "Can the ends and means of education be measured?" and "Do the presumed means in fact cause the ends, and, if so, to what extent or with what degree of effectiveness or productivity?" (Walberg, 1978).

Each of these major questions about education is important, though far from being completely answered. This seems particularly true as regards Follow Through as an education program ("means") designed to produce a variety of educational outcomes ("ends"). This paper specifically focuses on issues related to the evaluation of program implementation in Follow Through. Of necessity, some discussion of educational outcomes will be included. However, attention will be primarily focused on implementation as a measurement and evaluation notion in the national Follow Through program. Issues related to educational outcomes and their measurement and evaluation in Follow Through can be found elsewhere (e.g., House, Glass, Mclean, & Walker, 1978).

The first "planned" variation experiment in Follow Through has come

to an end nationwide. The results of research and evaluation efforts aimed at verifying the impact of various sponsor educational model treatments on children during the early school years have been presented in many sources. Reviews of this research by members of the professional education community have, at best, been mixed. Considerable variation in the effects of sponsors' educational models has been documented, with the implication that some models were more effective and educationally productive than others (Bock, Stebbins & Proper, 1977). Considerable variation in the magnitude of effects across sites receiving common sponsor treatments has also been observed. Since serious questions about the past effectiveness of the national Follow Through effort have been raised (and continue to be widely debated), there seems an obvious need to examine closely issues considered important to the success of a new series of research and evaluation efforts in Follow Through.

Past research and evaluation efforts in Follow Through have raised considerable controversy in both educational and political arenas, especially with regard to the effectiveness of sponsor models in bringing about positive growth in Follow Through children. The criticism can be made that the first planned variation experiment in Follow Through (FT) was not an "experiment" at all, but rather a conglomeration of post hoc data analyses attempting to support sponsor model effects. The National Institute of Education's (NIE's) desire to carefully plan for a new round of research efforts in Follow Through seems timely, particularly as regards the evaluation of program implementation. Newer conceptions of the measurement and evaluation of program implementation (when combined with the collective wisdom gleaned from our past experiences in Follow Through) can broaden our understanding of how innovative programs designed to serve disadvantaged children work.

A Rationale for Measuring Implementation in Follow Through

It is probably a fair judgment to say that the major emphasis of educational program evaluation models in the past has been on school and pupil-related outcomes. This seems particularly evident when one considers the national focus on academic learning and pupil achievement as measured by large-scale standardized testing, amidst the current era of heightened educational accountability. It might also be observed that models of educational productivity and policy analysis have developed contemporaneously with the concern for the evaluation of implementation of educational innovations. Thus, understanding issues concerned with the implementation of educational innovations targeted for disadvantaged populations (e.g., Title I, Head Start, Follow Through) has particular relevance at this time.

I find these observations interesting since models in educational research and evaluation conceptually follow those found in other more mature and exacting disciplines such as medicine and agriculture. The surgeon wants to know the recovery rate of patients receiving a particular surgical treatment when compared to other treatments or non-treated controls. Similarly, the agronomist is interested in crop yields in response to different soil treatments and weather conditions. Once causal relations are established in these disciplines, productive and cost effective treatments can be identified for desired outcomes and the policies for implementing important treatment aspects can be formulated. Finally, frameworks for evaluating whether important treatment events are being implemented can be developed for quality control.

Educational evaluation as a discipline is not as far advanced as either medicine or agronomy. Establishing causal relations between educational treatments and outcomes has not been as fruitful as similar attempts in other disciplines owing to many factors. Among the more obvious of these

are: 1) the imprecision of measurement methodologies; 2) the complexities of human behavior; 3) the complexities of education as a social system; 4) the logistical difficulties and impracticalities involved in carrying out "true" experiments in applied settings; and 5) the immaturity of the discipline itself.

In a recent article, Leinhardt (1980) concluded that educational programs (at least at the classroom level) should be evaluated in a two-staged process. First, educational treatment(s) should be directly measured and related to student outcomes in a causal fashion with explicitly stated models. Secondly, programs should be described in terms of the quantity and quality of treatment dimensions that they are observed to supply in natural educational settings. The first stage is one of providing the program developer with information about the degree to which various aspects of the innovation are being implemented. The second stage represents a further specification of implemented treatment(s) in terms of pupil outcomes. First stage, or implementation studies, according to Leinhardt (1980), "will not clarify why one set of approaches work better than another, nor are they likely to advance our understanding of the relationship between compensating instructional processes, though they may yield considerable information on dissemination and diffusion." If one of the current goals of research in education is to better understand the "means" (particularly at the classroom level) which produce important "ends" (pupil outcomes), then why study implementation at all?

In a landmark conceptual review, Fullan and Pomfret (1977) specified four major reasons why implementation should be studied in education. Each of these reasons seems important when viewing both past and future evaluations of the Follow Through program. First, is that "we do not know what has changed unless we attempt to conceptualize and measure it directly." Follow Through

is a program primarily based upon the external "change agent" philosophy with its current organizational arrangement of sponsors and sites. Since no attempts to systematically evaluate implementation of Follow Through program components (instruction, parent involvement, staff development and comprehensive services) either within or across sponsor models occurred at the program's beginning, little has been learned about these components and their relationship to program outcomes. Fullan and Pomfret's (1977) description of how innovations were conceptualized a decade ago seems quite characteristic of the history of the Follow Through program:

> "The assumption appears to have been that the move from the drawing board to the school or classroom was unproblematic, that the innovation would be used more or less as planned, and that the actual use would eventually correspond to planned or intended use. The whole area of implementation, what the innovation actually consists of in practice and why it develops as it does, was viewed as a 'black box' where innovations entering one side somehow produce the consequences emanating from the other." (p. 337)

It seems clear that future evaluations of Follow Through necessitate the measurement of implementation of key program components if the contents of Fullan and Pomfret's "black box" are to be better understood.

Secondly, it is important to evaluate implementation to "understand some of the reasons why so many educational changes fail to become established." I think most current Follow Through sponsors would agree that considerable variation in the extent to which sponsor models are established in associated sites exists. Most could even rank order their sites from "the most highly implemented program" to "the least implementation of all." It seems inevitable that some modification of Follow Through sponsor models will occur at the local district level. Explanations of the variation in the "fidelity" of implementation at the district or classroom level will continue to rest with opinion and authority in the absence of implementation data.

A third reason for studying implementation is "that the failure to do

so may result in implementation being ignored, or else being confused with other aspects of the change process . . . or even the confusing of the determinants of implementation with implementation itself." This reason highlights why levels of program implementation have often been inferred from pupil achievement and other outcomes in Follow Tnrough, rather than directly measured in their own right. Similarly, as Leinhardt (1980) has noted, studying uniform classroom process variables across Follow Through models may rank them in terms of degree of implementation (Kaskowitz & Stallings, 1975; Stallings, 1974), but the content of implementation may be masked.

A fourth reason for studying implementation according to Fullan and Pomfret (1977) is that "it may be too difficult to interpret learning outcomes and to relate these to possible determinants." This fourth reason is an important one. After 13 years of operation of national Follow Through, few facts have accumulated that demonstrate how implementation of Follow Through components either relate to or produces program outcomes. Classic cases in point are current Follow Through sites that have been nationally validated as "exemplary" early childhood education programs by the Joint Dissemination Review Panel. I am not aware of any sites that have been so validated on the basis of demonstrated relationships between measures of program implementation and program outcomes. I include here the Mathemagenic Activities Program (MAP) with which I am associated at the University of Georgia. All five school districts implementing the MAP educational model were recently (February, 1981) validated by the JDRP using math achievement of Follow Through children in the absence of program implementation data. Replicated achievement test gains with Follow Through eligible children are impressive and they can be readily evaluated in terms of their statistical and educational significance. Such gains become far more

important however, when they can be meaningfully related to program implementation data.

Given Fullan and Pomfret's (1977) rationale for measuring and studying implementation and the nature of the Follow Through program, developing measures of implementation of program components for future studies in Follow Through seems of high priority. A more detailed examination of implementation as an evaluation construct can perhaps elucidate its importance to the national Follow Through Program.

Implementation as an Evaluation Construct

From my viewpoint, implementation as a construct in educational evaluation can be considered from three perspectives: 1) the process perspective; 2) the independent variable perspective; and 3) the dependent variable perspective. Each of these has implications for the future study of implementation in Follow Through.

From the process perspective, an attempt is made during early stages of the innovation to describe key program components and their interworkings and relationships. Weak measurement models are permissible and often case study methodologies are used. The process perspective of implementation is probably most useful during the initial stages of development and "start-up" of an innovation when program planners are attempting to identify essential program elements. The development of a useable evaluation framework for national Follow Through which begins with a baseline description of key components of the program has been recently proposed by Wang and Ellett (1980). This approach identifies key program components through comprehensive description and survey methodology. The approach attempts to take into account the fact that Follow Through is a multidimensional program which has four key, generic components; instruction,

parent involvement, staff development, and comprehensive services. However, Follow Through models differ in their philosophical and theoretical orientations in the areas of early childhood development and education, as well as in the specific approaches they employ to address the four generic program components. Understanding common and unique ways in which Follow Through sponsors and sites address each component is a first step in evaluating program implementation.

Much has been written about what Follow Through is and should be. However, no systematic efforts are known to this author that have as a goal a comprehensive description of how program components are translated into practice across the variety of Follow Through models or school districts within models. If implementation as an evaluation construct is to be studied in future Follow Through research, arriving at a sound national program description seems a most logical place to begin.

From the independent variable perspective, implementation as an educational evaluation construct is considered to influence program outcomes. Stronger measurement models and methods are required here than at the program description level. The most detailed and comprehensive example of this implementation perspective to date is probably that described by Hall and Loucks (1977) and their notion of Levels of Use (LoU) of an innovation. It is interesting to note that levels of the LoU interview model were developed with the methodology alluded to above in the description of implementation from the process perspective.

According to Hall and Loucks (1977), many educational change and diffusion researchers (Fullan & Pomfret, 1977; Havelock, 1971; Rogers & Shoemaker, 1971) assume an innovation has been implemented once it is adopted, and the use of the innovation in the classroom or school remains essentially undocumented. The LoU interview methodology was developed to

measure 8 levels of use of an innovation: 1) Nonuse, 2) Orientation, 3) Preparation, 4) Mechanical Use, 5) Routine Use, 6) Refinement, 7) Integration, and 8) Renewal. Inter-rater reliabilities reported for the LoU (.87 to .96); the fact that it is considered generic and easily adapted to measuring many different innovations; and the practicalities of time of administration (20 minutes); suggest the LoU might be eventually useful as a gross index of the level of implementation of Follow Through components.

From the dependent variable perspective, implementation can be considered a program outcome in its own right. This notion implies that measures of implementation can be used as criteria against which the effects of key program features can be tested. For example, degree of implementation of a particular instructional program in Follow Through classrooms may show positive relationships to key model features such as indices of the quality of teacher aide training and the number of hours of inservice education received. During secondary stages of the establishment of innovations, measurements of degree or fidelity of implementation can assist in identifying program elements that are working and those that are not. High fidelity of implementation alone, however, should not be considered a terminal goal in Follow Through or in other educational programs.

In the evaluation of educational models at some midpoint in their development, fidelity of implementation might be conceptualized as a dependent (criterion) variable in the analyses. During later and more mature stages of model development and with more sophisticated data analysis strategies, fidelity of implementation could be one of a number of important independent variables included in an evaluation design using, for example, pupil achievement as a program outcome in the analyses. What is suggested here is the notion that innovations such as Follow Through go through stages of development as they move from the point of initial program installation, to

full implementation, to impact on key program outcomes.

Current Models for Measuring Program Implementation

To my knowledge, no comprehensive attempt has ever been undertaken to directly assess level of implementation of Follow Through's major components (instruction, parent involvement, staff development, and comprehensive services) and to relate such assessments to program outcomes. Indeed, this would be a formidable undertaking. Sponsors and sites emphasizing one or two of these components as a major program focus have undertaken implementation studies on what I consider to be a rather small scale.

Similarly, few studies can be found that have attempted to compare levels of implementation (in even one or two program components) across Follow Through models and then tie these implementation levels to program outcomes. The most frequently cited exception is probably the work of Stallings (1974). A later study of cross-site implementation that should be noted is the work of Leinhardt (1977). Large scale studies of achievement of Follow Through and non-Follow Through children have generally tended to infer implementation rather than directly measure it.

While no comprehensive studies and only a few small scale studies of implementation have been undertaken in Follow Through, several models with potential for assessing implementation in Follow Through have been developed. None of these models was originated with the inception of the Follow Through program. However, they each contain procedures and conceptual notions which can be adapted to selected aspects of Follow Through. Each also has apparent limitations.

An early (1969) evaluation model posited by Alkin required that data relevant to the extent of program implementation be collected. This approach was based on the idea of fostering full levels of implementation

before assessing program outcomes such as pupil achievement. Alkin's model has much in common with the "fidelity" of implementation approach. To postulate that program implementation can be examined without knowledge of characteristics of the program participants somewhat restricts the application of the model to Follow Through classrooms.

A second evaluation of implementation paradigm has as a primary focus direct, systematic classroom observation. This model is most closely associated with the work of Stallings (1974), Stallings (1975a), Leinhardt (1976), Leinhardt (1977), Leinhardt (1980), and Evans & Behrman (1977). An assumption of the direct observation paradigm for assessing implementation is that "key elements" of the innovation can be identified by program developers at an appropriate stage in the program's development. Classrooms in which these key elements are judged as being highly implemented are then evaluated in view of educational outcomes (such as pupil achievement). Systematic classroom observation has an intuitive appeal as a preferred method for measuring implementation. However, observational data are sometimes difficult and expensive to collect and observational methodologies are frequently misunderstood in terms of sources of error and dependability for making decisions (McGaw, Wardrop & Bunda, 1972; Capie, Tobin, Ellett, & Johnson, 1981). Even with these methodological and logistical difficulties, systematic observation of the operation of key program variables seems the preferred measurement technology for evaluating program implementation.

The "levels of use" of an innovation model put forth by Hall and Loucks (1977), and Loucks, Newlove & Hall (1975) has been previously described. The structured interview methodology has promise as an easily adaptable approach to assessing implementation. The LoU is structured to provide a set of descriptions of behavior of individuals from the

time before any knowledge of an innovation exists until the innovation reaches a level of possible expansion and revision . . . going from relative immaturity to a mature state.

A relatively new approach to measuring program implementation has been put forth by Churchman (1979). His approach assumes that it is virtually impossible to accomplish complete implementation of any educational program because those involved in the innovation (teachers, pupils and others) influence the level of implementation achieved. Churchman's approach involves the notion of relating variations in teachers' adaptations to curricula to variations in learner outcomes using structural equations.

Churchman's model for assessing implementation has been criticized in at least two ways by Revicki and Rubin (1980). First, the model fails to capture the structural integrity of innovative programs. Since educational programs tend to be related to various conceptual and theoretical models, program features should not be viewed as random collections of variables which can be simply ignored, varied or implemented by program participants. Secondly Churchman makes no clear recommendation concerning how data should be collected.

Newfield (1979) has recently developed a method of assessing program implementation that involves repeated measures of key program features. Multiple-matrix sampling is used in the data collection procedures to insure reliable measurements and to ease logistical difficulties. The use of the multiple-matrix sampling strategy also allows for measurements of a large number of program variables with inputs from large numbers of program participants. According to Revicki and Rubin (1980), the Newfield implementation model is difficult to use for formative evaluation purposes because of the complexity of the sampling and data analysis procedures. Additionally, the primary data collection method of self-reporting raises validity and reliability questions.

In reviewing current models for evaluating implementation of educational programs and "treatments" one point seems quite clear. The entire notion of evaluating implementation is still in its infancy. This seems particularly the case if one was to use any of the current models to evaluate educational program implementation in national Follow Through. With this concern, each of the models described above has considerable shortcomings, though each also has quality points as well.

First, none of the models is comprehensive enough in terms of variables measured to evaluate the plethora of sponsor model "treatments" in the national Follow Through program. Secondly, none is comprehensive nor flexible enough in terms of data collection and analysis methodology. Thirdly, none were specifically developed with Follow Through goals in mind (the possible exceptions being the work of Stallings and Leinhardt and their colleagues).

Because of the comprehensive nature of Follow Through as an educational innovation in early childhood education, new and more comprehensive models for evaluating program implementation are needed. This seems particularly the case across sponsors. For example, most available methods for measuring program implementation are derived from studies using classroom or school-related variables. Follow Through is a comprehensive program in terms of emphasis. The University of North Carolina, for example, sponsors a Parent Education Follow Through program which has an educational philosophy and focus quite different from the Mathemagenic Activities program at the University of Georgia, the Behavioral Analysis Model at the University of Kansa, or other sponsor models. Any system of methods developed to evaluate implementation across Follow Through models must consider this fact. Most assuredly, variation in Follow Through sponsor treatments and in the communities with which they work is a reality. Similar viewpoints have been rather

widely expressed concerning the nature of program "outcomes" for national Follow Through.

The problem delineated here is one of developing a program implementation system around key classes of national Follow Through program variables (instruction, parent involvement, staff development and comprehensive services) which is flexible enough to be used by the wide variety of sponsor models to examine program effects. This presents quite a formidable task given the current level of development of models for measuring implementation within Follow Through. But, I don't believe the task is insurmountable if it is given philosophical priority and necessary human and financial resources. Certainly we have the technology in the electronic age of high speed computers and we know enough about research design, educational and psychological measurement, and data analysis strategies and techniques. General procedures and measurement examples for undertaking such a large effort are discussed later in this paper. What kinds of issues and questions pervade the development of a systematic approach to evaluating implementation in Follow Through?

Implementation and/or Treatment Measurement?

There are many conceptual and methodological issues that need to be resolved before an adequate system for evaluating implementation in the national Follow Through program can be designed. First is to find an acceptable conception of implementation as an evaluation construct.

I previously expressed the viewpoint that implementation is developmental in the sense that it can be considered to go from a relatively immature state to a rather mature one. This view seems consonant with that expressed by Hall and Loucks (1977) and many others. From the developmental perspective implementation is a sociological, school organizational phenomenon that deserves study in its own right and generates its own set of important

questions about people, their goals, their roles, and their educational values. What are the key human and organizational factors that lead to the installation, maintenance, maturity and success of a Follow Through program? Are there unique factors which serve to inhibit or facilitate Follow Through implementation at the local school level? What changes in school structure, organization and management encourage or discourage adoption of a sponsor model as an educational treatment? What initial, operational and long-term changes in school priorities are made as a result of Follow Through program implementation? What are the critical features of implementation that lead to the most rapid change in schools? Are their distinct stages in the implementation process that are generic across different Follow Through sponsor models? How does specific training of school personnel assist the pace at which Follow Through programs are implemented? At what level of implementation can a Follow Through program continue without external assistance from sponsors acting as "change agents?"

This list of questions is by no means complete. It simply represents a sampling of the kinds of questions those desiring to study implementation as a phenomenon in its own right would have us ask. An important issue in this regard is whether energies and monies should be expended to understand implementation as a process of program development and change; or whether it should be further studied with an eye toward evaluating program effectiveness and educational productivity.

A second major issue is how implementation fits on the measurement continuum and what decisions can and should be made with implementation data. I share the view that implementation is at one end of the measurement continuum and "treatment" at the other. A view similar to this one has been recently expressed by Leinhardt (1980). Implementation as a program evaluation construct allows for "weaker" and more global measurement indices.

Judgments about the degree and fidelity of program implementation derived from interview data and subjective judgments, for example, fit this view. Such measures might be useful in answering broad questions about Follow Through such as: At what stage of development is the program? Is the program working? Which program components are operating? Which program components are still to be developed?

At the other end of the continuum is educational "treatment" which requires greater measurement precision and more detailed specification of program variables. Measuring treatments (as opposed to implementation) in Follow Through leads to other sorts of questions. Which aspects of the classroom instructional process contribute most to pupil learning? How does the variation in time allocated to instruction impact on program outcomes across Follow Through school districts? Which instructional procedures and strategies impact on pupil perceptions of the learning environment? Which aspects of aide training have the greatest impact on the organization and operation of Follow Through classrooms? How are program evaluation data being translated into teaching strategies in Follow Through classrooms? What impact are parent involvement programs having on pupil academic engagement at home? Clearly these questions are also important to the future production of knowledge in Follow Through.

Conceiving of implementation at one end of the measurement continuum and educational treatment at the other is, agreeably somewhat arbitrary. However, I think it helps clarify which kinds of questions can be answered and in which ways. One might postulate as Leinhardt (1980) has done, that if all important educational processes in an educational model are specified and their contribution to the production of outcomes understood (i.e., amounts of outcome variance explained), then the addition of gross treatment data adds little to our understanding. In Leinhardt's words, "the weight for

T (treatment index) should be insignificant in the presence of the P's (specific measures of process) if the P's are capturing the important treatment differences." Similarly, the weight for I (implementation score) should be insignificant in the presence of the T's (treatment measures) if the T's are capturing the important model, school, classroom or child differences.

Again, a clear distinction between implementation and treatment evaluation is difficult to make. But, it seems certain that a measurement focus on implementation alone in future studies in Follow Through can mask important educational treatments and the production of useable knowledge. For example, knowledge such as, "the program is being implemented in Follow Through classrooms with a high degree" does not seem nearly so important to me as the knowledge that "the greater skill of Follow Through teachers in classroom management this year than last has eventuated in more puil academic engagement." The first bit of knowledge is possibly useful in making a global judgment as to whether the program is being implemented. The second bit of knowledge is a specific statement about educational treatments eventuating in pupil-related outcomes.

Similarly, measuring the number of parents involved in Follow Through classrooms in an academic year might be useful information from the evaluation of implementation perspective. However, understanding the impact of parental involvement in the classroom and the manner in which such involvement leads parents to change the educational quality of the home environment seems far more important.

Viewing national Follow Through as a program that varies among models, schools, and classrooms in terms of degree of implementation is one perspective and an important one. Understanding Follow Through as a series of educational treatments seems to be another. Both have analogues in designing a comprehensive evaluation/experimentation plan for future studies in Follow Through.

Defining the "What" of Implementation

The development of a framework for evaluating both implementation and the variety of educational treatments in Follow Through must begin with a systematic and comprehensive description of key program components. What should be implemented in Follow Through is a prior question to the development of implementation measures. Any comprehensive description of Follow Through as an educational innovation must take into account the program's multi-dimensionality. That is, it serves as a resource for early childhood development and education through four program components: instruction, parent involvement, staff development, and comprehensive services. The program emphasizes both the delivery and documentation of effective service related to these four components and the production of knowledge that can be used to better understand the nature of the program as a series of educational treatments, and the influence of these treatments on outcomes. Thus, Follow Through can be viewed as a program designed to provide effective service to disadvantaged children, their families and communities, as well as a laboratory for knowledge production activities aimed at studying and assessing innovative ways of providing effective service.

With its current organization of sponsor model/site associations, the "planned variation" nature of the national Follow Through program must also be considered in the description of the program and in the design of an overall evaluation framework. Follow Through includes a diversity of educational models, each of which must be respected in any attempt to define and/or evaluate effective service. The purpose of each educational model in Follow Through is to demonstrate viable alternatives for achieving national Follow Through's overall goals. However, the models differ in their philosophical and theoretical orientation to early childhood development and education as well as in the specific approaches they employ to provide services

that are considered effective in responsively and effectively meeting the needs of Follow Through children and their families.

While the contribution of each of the models represented in Follow Through must be recognized in any description and/or evaluation of implementation in the overall program, providing model-specific information alone seems insufficient. From the decision-making and policy analysis perspectives, implementation data on the program as a whole seems desirable. Thus, efforts to evaluate implementation in Follow Through must consider the program as a whole and must at the same time recognize the individuality of each sponsor model as well. Is it possible to develop measures to evaluate implementation of the national Follow Through program given its comprehensive nature while maintaining the individuality of sponsor models? And if so, how can such an effort proceed?

At least two kinds of information are needed. First, each Follow Through sponsor must provide general categories of information about the four program areas: instruction, parent involvement, staff development, and comprehensive services. This information would be in response to what is being implemented in each program component. Secondly, information is needed concerning the manner in which each model is being implemented: the how of implementation.

Wang and Ellett (1980) have provided a general description of a methodology useful in developing a system for evaluating program implementation in Follow Through. During Phase I of their proposed effort each sponsor would be asked to provide selected information on their Follow Through model. Such information would include details about the manner in which the model addresses each of the four major Follow Through program components. While written description of most sponsor models already exist, Wang and Ellett (1980) propose a systematic data collection plan which can possibly contribute to formulating a description of the Follow Through program on a whole. Data collection

methodologies would include formal surveys of and selected interviews with sponsor and site staff, analyses of existing sponsor and site records, and scheduled and random observations of program activities related to each of the four major components of national Follow Through. Such information would be used to define "critical dimensions" of each of the Follow Through components.

Critical dimensions and sets of associated scaled descriptors would also be developed to further cast the four program components into a measurable framework. Where appropriate, performance indicators for the critical dimensions of program components would utilize existing information such as that identified by Applied Management Sciences (1979) and model sponsors (e.g., Wang, 1980; Ellett, et al., 1980). A process model would then be used to synthesize the comprehensive performance indicators for each critical dimension and a generic framework for evaluating Follow Through program implementation would be structured. The scaled descriptors would provide a measure of the implementation of the four program components by which their identified critical characteristics can be monitored and evaluated. The evaluation framework would thus be "generic" in terms of the what of implementation, but flexible in nature in order to adaptively accommodate the diversity of Follow Through models.

Examples of "critical dimensions" of two Follow Through program components, along with associated performance indicators and scoreable scaled descriptors proposed by Wang and Ellett (1980) follow. The measurement methodology, depicted in these examples is currently being used in Georgia to reliably assess beginning teachers' competencies for initial certification (Capie, Tobin, Ellett, and Johnson, 1981).

Developing instrumentation for the evaluation of implementation of Follow Through components is indeed an ambitious undertaking. However, the what of implementation in Follow Through, or in any other educational program, precedes its measurement and evaluation.

FOLLOW THROUGH MAJOR COMPONENT: INSTRUCTION

Follow Through Critical Dimension: Teacher Classroom Performance in "Communicating with Learners"

Performance Indicator: Teacher clarifies directions and explanations when learners misunderstand content.

Scale of Scoreable Descriptors:

1. Discourages learners when they seek clarification on directions or explanations.

2. Ignores learners when they seek clarification, directions, or explanations.

3. Restates original communication in nearly the same words if learners do not understand.

4. Gives directions or explanations using different words and ideas when learners do not understand.

5. In addition to the items in number 4 above, the teacher attempts to identify areas of misunderstanding and to restate communication before learners ask.

OR

No misunderstanding by learners was evident during the lesson.

Comment

The sample above is just one example of a possible item to be included in a classroom observation system to assess teacher classroom performance as only one critical dimension of the FT component of Instruction. The Instruction component undoubtedly has many other facets to be measured besides teacher performance, such as classroom environment characteristics, instructional planning, pupil behavior, instructional materials, etc. The scale of descriptors is arranged hierarchically from a low rating of 1 to a high rating of 5.

FOLLOW THROUGH MAJOR COMPONENT: PARENT INVOLVEMENT

Follow Through Critical Dimension: Parent Participation in Classroom Activities

Performance Indicator: Parents are actively involved with classroom instruction of Follow Through children.

Scale of Scoreable Descriptors:

1. No records, observations, or information are evident to verify parent involvement in classroom instruction.

2. Formal discussions with site personnel suggest some parent involvement in classroom instruction but no documentation exists to verify this information.

3. Formal discussion with site personnel suggests moderate degrees of parent involvement in classroom instruction with some documentation available to verify this information.

4. Formal discussions and records indicate appropriate amounts of parent involvement in classroom instructional activities.

5. In addition to the information in 4 above, formal discussions, records and observations indicate a high degree of parent involvement in classroom instructional activities. This amount of involvement is beyond that considered minimally essential for implementation of the sponsor model.

Comment

The sample above is just one example of a possible item to be included in a data collection instrument to assess parent participation in classroom activites as only one critical dimension of the FT component of Parent Involvement. This component undoubtedly has many other facets to be measured such as parents working in groups, parents working in the community, parents participating in continuing education, etc. The scale of descriptors is arranged hierarchically from a low rating of 1 to a high rating of 5.

A first objective in future efforts to study program implementation in Follow Through should be to provide an empirically based description of what the program is and is not. Only then, can a system of measurements and documentation be developed to assess program implementation. If these tasks could be accomplished in a manner which incorporates essential program components, while protecting the individuality of sponsor models, then Follow Through would have a measurement system which could be used to collect data for program decision making, future policy analysis, and knowledge production as well. Without such a system, the "black box" view of program evaluation (Fullan and Pomfret, 1977) will continue in Follow Through with degree of program implementation remaining an inference derived from program outcome data alone.

Implementation, Treatment, and Outcome Relations in Follow Through: Data Analysis Strategies

A host of data collection and analysis strategies have been used in past "national" and sponsor-initiated studies in Follow Through. However, the most frequently used data analysis model has been the comparison of "treated" Follow Through children to their "non-treated" counterparts. Continued comparative studies of this type may not be the most fruitful approach for future studies of implementation and treatment effects in Follow Through. There are several reasons why this is so.

First, there has often been confounding of Follow Through and non-Follow Through children with eligibility requirements (income level) and their actual treatment in Follow Through classrooms. Pure Follow Through and non-Follow Through classrooms (with appropriate income eligibility controls) have been difficult to maintain in past research studies. Some sponsors have maintained Follow Through classrooms consisting of only Follow

Through eligible children, and non-Follow Through classrooms consisting of Follow Through eligible but non-treated controls. Other sponsors have had difficulty in controlling eligibility and treatment, and as a result, these limitations have adversely influenced potential comparisons between Follow Through eligible and non-Follow Through eligible children. Most Follow Through classrooms today consist of income eligible and non-income eligible children. Given the practical constraints of local school organization, these facts seem to obviate the utility of the Follow Through to non-Follow Through comparison model for future research and evaluation efforts.

Similarly, in the spirit of "proliferating what is good about Follow Through," non-Follow Through classrooms are often treated like their Follow Through counterparts because successful teaching practices are adopted by non-Follow Through teachers. This fact somewhat obviates comparisons of Follow Through to non-Follow Through classrooms within schools.

Secondly, longitudinal studies of Follow Through children using a large number of process and outcome measures greatly increase the cost of data collection, processing, and analysis and dramatically inflate program costs per child. Similarly, collecting large amounts of data on hosts of program variables usually necessitates variable combining and reduction strategies when data analyses are undertaken.

Thirdly, as indicated earlier, simple comparisons between Follow Through and non-Follow Through children to demonstrate that a given Follow Through model "works" or is effective are not sufficient in terms of what the data says to potential adopters. Far more important is the explanation of these differences in terms of program implementation and treatment measures. In future studies in Follow Through it would seem important to sacrifice massive collection of data on non-Follow Through children for a more intense concentration on the measurement of key program processes and implementation characteristics.

Some attention in future Follow Through studies should be given to studying "intermediate" model effects (Ellett, Hawn, Pool, and Smock, 1979). These effects are considered important for understanding long-term program outcomes. Figure 1 presents a summary of interrelationships among key classes of variables considered important for future research and evaluation studies in Follow Through. Implementation factors are conceptualized as having a primary impact on a class of pupil-related "intermediate" or variables intervening between program implementation and outcomes. Program outcomes such as pupil achievement are considered to be affected by background characteristics (such as the quality of the home environment), program implementation factors, and intermediate variables as well. In this model, both individual child characteristics and level of model implementation are conceptualized as determiners of intermediate variables, regardless of a particular sponsor model's program or philosophy. Intervening (intermediate) variables in turn, are conceptualized as affecting achievement with antecedents (background variables) having an independent, direct influence (broken arrows) on achievement. Undefined factors can also affect states of intervening variables in the model as well as subsequent learner outcomes. The model suggests the importance in future Follow Through studies of examining not only the relative contribution of known achievement correlates to Follow Through children's growth, but, in addition, to undertake small studies of undefined factors and the contribution of program implementation.

The model also suggests that "causal modeling" in future Follow Through studies should be the primary analytic tool rather than a continued comparison of Follow Through to non-Follow Through children in terms of program outcomes. Relationships between classes of variables <u>within</u> and <u>across</u> sponsor models could be more formally structured through deriving a variety of functional equations of the form:

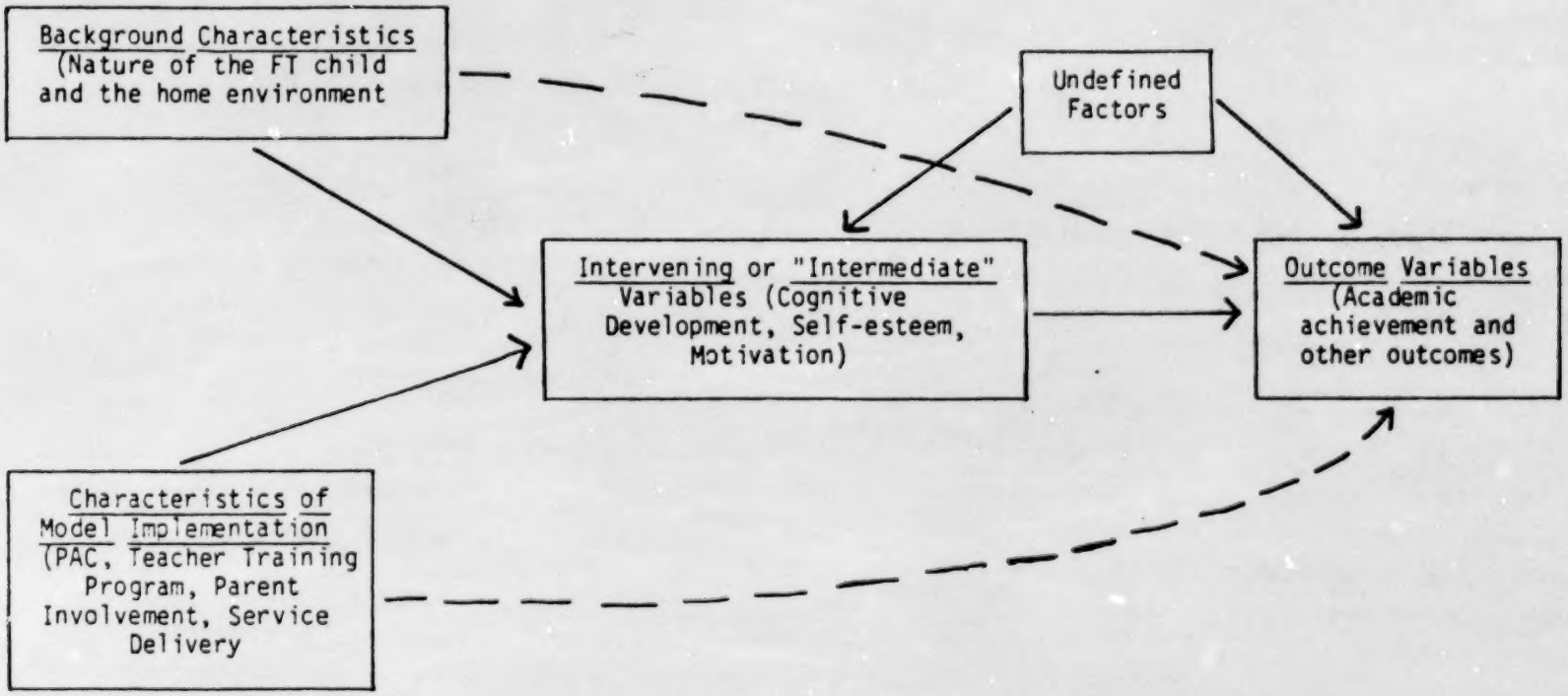
Background Characteristics
(Nature of the FT child and the home environment
Undefined Factors
Intervening or "Intermediate" Variables (Cognitive Development, Self-esteem, Motivation)
Outcome Variables
(Academic achievement and other outcomes)
Characteristics of Model Implementation
(PAC, Teacher Training Program, Parent Involvement, Service Delivery

Figure 1. Summary of Interrelationships Among Variable Classes for Future Studies in Follow Through

$$a = f(b, m, i, u)$$

where,

a = pupil achievement

b = background characteristics (including pretest achievement)

m = model implementation

i = intermediate or postulated intervening variables

u = a collection of undefined factors

In developing such functional equations, emphasis is given to the "relative contribution" of the variables investigated to progress in Follow Through children, rather than to a comparison of Follow Through and non-Follow Through samples on a single outcome measure. The model is more flexible and comprehensive than past analytic models used in Follow Through, and it has the advantage of initiating explanations of Follow Through effects in terms of child characteristics and model implementation factors. In addition, if common metrics could be established for broad classes of variables (e.g., model implementation) as previously suggested, an analysis of the contribution of these variable classes to productivity in Follow Through children could be made.

There is no reason to believe that independent variables in the data analysis model act autonomously in determining levels of specified outcomes however. In fact, interactions between "production" variables would be expected and relationships would not necessarily be linear. An exaplanation and examples of co-linearity of variables in such functions as that proposed above have been discussed by Walberg (1978) and others.

An application of educational productivity models to the comprehensive evaluation of implementation and treatment effects in Follow Through requires: 1) a design which allows for systematic collection of program implementation data; and 2) statistical tools for sorting out the contribution of program

implementation to program outcomes in the presence of other known outcome variable correlates. Statistical tools to test such models are available and have been extensively used in econometric analyses (Hanushek, 1978; Lau, 1977). However, a comprehensive and systematic approach to assessing implementation of Follow Through's four key components (instruction, parent involvement, staff development, and comprehensive services) is not yet available.

If we are to move forward in our understanding of program implementation effects in Follow Through, development of implementation measures should receive a first priority. This will be a difficult and expensive task. However, it seems the only way in which we can respond to a critical question about Follow Through "Do the presumed means in fact cause the ends, and, if so, to what extent or with what degree of effectiveness or productivity?" Once we understand the means of Follow Through and their contribution to the ends of Follow Through, resources can be allocated accordingly.

References

Alkin, M.C. Evaluation theory development. Evaluation Comment. 1969, 2(1).

Applied Management Sciences. Development of a performance monitoring system for Follow Through local service projects. September, 1979.

Bock, G., Stebbins, L.B., & Proper, E.C. Education as experimentation: A planned variation model. In Effects of Follow Through Models (Vol. 4-b). Cambridge, Mass.: Abt Associates, Inc., 1977. (Also issued by the U.S. Office of Education as National evaluation: Detailed effects, Volume II-B of The Follow Through Planned Variation Experiment series).

Capie, W., Tobin, K., Ellett, C.D. & Johnson, C.E. The dependability of job performance rating scales for making classification decisions. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

Churchman, D. A new approach to evaluating the implementation of innovative educational programs. Educational Technology, 1979, 25-28.

Ellett, C.D., Hawn, H.C., Pool, K., Des Jardines, L. The Mathemagenic Activities Program Implementation Assessment Instrument. Athens: College of Education, University of Georgia, 1980.

Ellett, C.D., Hawn, H.C., Pool, K. & Smock, C. Planning Information Study For Future Follow Through Experiments: Task I Response - General Issues. Athens: College of Education, University of Georgia, 1979.

Evans, W. & Behrman, E. Strategy for evaluating curriculum implementation. Curriculum Studies, 1977, 9(1), 75-80.

Fullan, M. & Pomfret, A. Research on curriculum and instruction implementation. Review of Educational Research, 1977, 47(2), 335-397.

Hall, G. & Loucks, S. A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 1977, 14(3), 263-276.

Hanushek, E.A. A reader's guide to educational production functions. Paper prepared for the NIE National Invitational Conference on School Organization and Effects, San Diego, 1978.

Havelock, R.G. Planning for innovation through dissemination and utilization of knowledge. Ann Arbor: Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, The University of Michigan, 1971.

House, E.R., Glass, G.V., McLean, L. & Walker, D. No simple answer: Critique of the Follow Through evaluation. Harvard Educational Review, 1978, 48(2), 128-160.

Kaskowitz, D. & Stallings, J. An assessment of program implementation in project Follow Through. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., 1975.

Lau, L.J. Educational production functions. Palo Alto: Paper prepared for the National Institute of Education, 1977.

Leinhardt, G. Observation as a tool for the evaluation of implementation. Instructional Science, 1976, 5, 343-364.

Leinhardt, G. Modeling and measuring educational treatment. Review of Educational Research, 1980, 50(3), 393-420.

Loucks, S., Newlove, B. & Hall, G. Measuring levels of the use of the innovation: a manual for trainers, interviewers and raters. Research and Development Center for Teacher Education, University of Texas, Austin, 1975.

McGaw, B., Wardrop, J.L. & Bunda, M. Classroom observation schemes: Where are the errors? American Educational Research Journal, 1972, 9(1), 13-27.

Newfield, J. Measuring the degree of program implementation. Manuscript, Department of Curriculum and Supervision, University of Gerogia, Athens, 1979.

Revicki, D. & Rubin, R. Models for measuring program implementation: A review and critique. Paper presented at the annual meeting of the American Educational Research Association, Boston, April, 1980.

Rogers, E.M. & Shoemaker, F.F. Communication of innovations. New York: Free Press, 1971.

Stallings, J. An implementation study of seven Follow Through models for education. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1974.

Stallings, J.A. A study of implementation in seven Follow Through educational models and how instructional processes relate to child outcomes. Paper presented at Conference on Research on Teacher Effects: An examination by Policy Makers and Researchers, Austin, 1975.

Stallings, J.A. Implementation and child effects of teaching practices in Follow Through classrooms. Monographs for the Society for Research in Child Development, 1975a, 40, 1-119.

Walberg, Herbert J. A theory of educational productivity. Invited address presented at the annual meeting of the Georgia Educational Research Association, Atlanta, 1978.

Wang, M.C. Proposal for continuation of implementation and evaluation activities under project Follow Through. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1980.

Wang, M. & Ellett, C.D. A data-based approach to providing a description of and evaluation framework for the national Follow Through program. Research proposal submitted to USOE, August, 1980.